ARTICLE IN PRESS

Economic Modelling xxx (2018) 1–16

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Moment-based tests for random effects in the two-way error component model with unbalanced panels

Jianhong Wu[a,*], Guodong Li[b], Qiang Xia[c]

[a] College of Mathematics and Science, Shanghai Normal University, Shanghai 200234, China
[b] Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong
[c] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

ARTICLE INFO

JEL subject classification:
C12
C23

Keywords:
Moment estimation
Random effects
Test for random effects
Two-way error component model
Unbalanced panels

ABSTRACT

This paper extends Wu and Li (2014)'s moment-based tests for random effects to the case with unbalanced panel data. Based on the difference of variance estimators of the idiosyncratic errors at different robust levels, two statistics are constructed to test for the existence of individual and time effects, respectively. Some variants of the two statistics and joint tests for both the two effects are also discussed. Their asymptotic properties are obtained under some mild conditions. It is shown that the tests retain the desired properties observed in the balanced panel data case. Monte Carlo simulation experiments and a real data analysis are carried out for illustration.

1. Introduction

In the econometric analysis of panel data, one mainly focuses on the case with balanced data (e.g. Baltagi, 2008). In practice, however, it is common to encounter missing observations in the collected data set. For example, in labor economics, some data on individual income may be dropped out after some time periods due to retirement. Throughout this paper, the considered panel data sets can be allowed to be unbalanced, that is, some data may be not observed in some time periods for some individuals. Actually, most panels encountered in practice are of the unbalanced kinds (see, e.g. Baltagi, 2008; Baltagi and Song, 2006). Statistical modeling analysis for unbalanced panels has not received the attention that it deserves. Note that, misspecification of the existence of random effects in the error component will lead to seriously biased standard errors and even inefficient statistical inference, which is completely similar to that of balanced panels. So, it is interesting to test for random effects in the regression model with unbalanced panels.

Till now, there are some relevant literature on testing for the existence of random effects in the error component regression model with unbalanced panels. In the following we give a simple review for some main literature. Baltagi and Li (1990) extended Breusch and Pagan (1980)'s LM tests to the error component model with unbalanced panels. Since the variance cannot be negative, the two-sided alternative hypotheses seem to be unreasonable. Moulton (1987) extended the uniformly mostly powerful tests (UMPT) of Honda (1985) to the unbalanced one-way error component model and illustrated this method by a hedonic housing price unbalanced panel data model. However, as Moulton and Randolph (1989) argued, using the asymptotic critical values for the test of Moulton (1987) can lead to incorrect inference, especially when there is high correlations among regressors or the number of regressors is very large. And then Moulton and Randolph (1989) suggested a standardized lagrange multiplier (SLM) test which had better critical value approximations. Some test statistics were similarly suggested for time effects, see, e.g., Baltagi et al. (1998), Honda (1985) and Moulton and Randolph (1989). However, these tests are based on the one-way error component models, i.e. the null hypothesis corresponds to the case without any effects, and then the sizes may be distorted due to the presence of the time (individual) effect when the individual (time) effect is tested. Besides, although these LM-based methods have simple forms, all of them require the assumption of normality of idiosyncratic errors and independence among regressors, random effects and idiosyncratic errors, which can not be guaranteed in practice. Wu and Li (2014) proposed several

* Corresponding author.
E-mail addresses: wujianhong@shnu.edu.cn (J. Wu), wjhstat1@163.com (J. Wu).

https://doi.org/10.1016/j.econmod.2018.05.003
Received 26 November 2016; Received in revised form 28 February 2018; Accepted 3 May 2018
Available online XXX

moment-based test statistics for the existence of random effects, which are shown to have many desired properties such as the simplicity, the robustness to the distribution assumptions and the possible dependency among regressors, random effects and idiosyncratic errors. However, their methods are only available for the case with balanced panels.

In this paper, we focus on hypothesis test for the existence of individual and time effects in the two-way error component model with unbalanced panels. Specifically, we extend Wu and Li (2014)'s moment-based test methods to the case with unbalanced panels. In Section 2, we outline the different forms of the original model and obtain a robust estimator of parameter coefficients β which is asymptotically normally distributed under some regular conditions. In Section 3, we construct the test for individual effect. We first derive two estimators of variance of the idiosyncratic error. One is the robust estimator which is consistent no matter of the existence of individual and time effects, and another one is consistent when the individual effect does not exist while inconsistent under the presence of individual effect. Based on the difference of the two estimators, we construct the test statistic for individual effects, which can be shown to asymptotically normally distributed. And we can show that our test statistic is more powerful than the traditional ANOVA F test when the regressors are correlated with the individual effect. In Section 4, we use the same method to construct statistics for testing time effect and study their asymptotic properties. In Section 5, we construct several joint test statistics for both the two random effects and study their asymptotic properties. It is shown that the tests retain the desired properties observed in the balanced panel data case. Monte Carlo simulations are given in Section 6. Section 7 applies our methods to a real data example. Section 8 gives some conclusions and discussions. All proofs are provided in the Appendix.

For the sake of statements, we first introduce some notation as follows. We denote by $A\prime$ the transpose of matrix A, by A^{-1} the inverse of matrix A, and by $\|A\| = [\text{tr}(A'A)]^{\frac{1}{2}}$ the norm of matrix A. $A \otimes B$ is the Kronecker product of matrices A and B, and $diag_L(A_l)$ is a block diagonal matrix with the diagonal elements A_1, A_2,…, A_L. The symbol "$\Rightarrow$" stands for weak convergence. $a_n = o_p(b_n)$ means that a_n/b_n converges to zero in probability, and $a_n = O_p(b_n)$ means that a_n/b_n is bounded in probability. $\mathbb{E}X$ or $\mathbb{E}(X)$ stands for the mathematical expectation of random variable X.

2. Model and notations

Consider the following two-way error component regression model with panel data,

$$y_{it} = \alpha + X_{it}'\beta + u_{it}, \quad u_{it} = \mu_i + \eta_t + v_{it}, \tag{1}$$

where α is a scalar, $X_{it} = (X_{it,1}, X_{it,2}, \ldots, X_{it,K})'$ is the it-th observation on K observable regressors, β is the vector of coefficients of the regressors, and u_{it} is the error component including the idiosyncratic errors v_{it} and two random effects (μ_i and η_t). The random effects μ_i and η_t are used to capture the heterogeneity of individual and time periods, respectively. Further, the individual effect μ_i is assumed to be independent and identically distributed with mean zero and finite variance σ_μ^2, and the idiosyncratic error v_{it} is assumed to be independent and identically distributed with mean zero and finite variance σ_v^2.

In the following, we assume that there are L disjoint subsets $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_L$ of $\{1, 2,\ldots, n\}$ such that the observed time periods are identical for each $i \in \mathcal{N}_l$ with $l = 1, 2,\ldots, L$. The subjects with individuals in $\mathcal{N}_l$ form a balanced panel data set with the same time periods (denoted by $t_{l,1}, t_{l,2}, \ldots, t_{l,T_l}$). And we denote the number of individuals in group $\mathcal{N}_l$ by n_l. For each group $\mathcal{N}_l$, model (1) can be rewritten into the vector form as

$$\mathbf{y}_{l_i} = \alpha \boldsymbol{\iota}_{T_l} + \mathbf{X}_{l_i}\beta + \mu_{l_i}\boldsymbol{\iota}_{T_l} + \boldsymbol{\eta}_l + \mathbf{v}_{l_i}, \quad l_i \in \mathcal{N}_l, i = 1, 2, \ldots, n_l, \tag{2}$$

where $\boldsymbol{\iota}_k$ is a vector of ones with dimension k, $\mathbf{y}_{l_i} = (y_{l_i,t_{l,1}}, y_{l_i,t_{l,2}}, \ldots, y_{l_i,t_{l,T_l}})'$, $\mathbf{X}_{l_i} = (X_{l_i,t_{l,1}}, X_{l_i,t_{l,2}}, \ldots, X_{l_i,t_{l,T_l}})'$, $\boldsymbol{\eta}_l = (\eta_{t_{l,1}}, \eta_{t_{l,2}}, \ldots, \eta_{t_{l,T_l}})'$, and $\mathbf{v}_{l_i}$ is defined similarly.

For each group $\mathcal{N}_l$, we first eliminate the time effect by centering each term in model (2),

$$\widetilde{\mathbf{y}}_{l_i} = \widetilde{\mathbf{X}}_{l_i}\beta + \widetilde{\mu}_{l_i}\boldsymbol{\iota}_{T_l} + \widetilde{\mathbf{v}}_{l_i}, \quad l_i \in \mathcal{N}_l, i = 1, 2, \ldots, n_l, \tag{3}$$

where $\widetilde{\mathbf{y}}_{l_i} = \mathbf{y}_{l_i} - \frac{1}{n_l}\sum_{i=1}^{n_l}\mathbf{y}_{l_i}$, $\widetilde{\mathbf{X}}_{l_i}$, $\widetilde{\mu}_{l_i}$ and $\widetilde{\mathbf{v}}_{l_i}$ are defined similarly. By the knowledges on Algebra, we can find a matrix Q_{T_l} such that $(\frac{\boldsymbol{\iota}_{T_l}}{\sqrt{T_l}}, Q_{T_l})$ is a $T_l \times T_l$ orthogonal matrix (e.g. Wu and Li, 2014). Premultiplying model (3) with the matrix Q_{T_l}' yields,

$$Q_{T_l}'\widetilde{\mathbf{y}}_{l_i} = Q_{T_l}'\widetilde{\mathbf{X}}_{l_i}\beta + Q_{T_l}'\widetilde{\mathbf{v}}_{l_i}, \quad l_i \in \mathcal{N}_l, i = 1, 2, \ldots, n_l, \tag{4}$$

since $Q_{T_l}'\boldsymbol{\iota}_{T_l} = 0$.

Model (1) can be rewritten into the matrix form by stacking the observation of the L groups as

$$\mathbf{y} = \alpha\boldsymbol{\iota}_N + \mathbf{X}\beta + Z_\mu\boldsymbol{\mu} + Z_\eta\boldsymbol{\eta} + \boldsymbol{v},$$

where $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_L')'$ with $\mathbf{y}_l = (\mathbf{y}_{l_1}', \mathbf{y}_{l_2}', \ldots, \mathbf{y}_{l_{n_l}}')'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_L)'$ with $\boldsymbol{\mu}_l = (\mu_{l_1}, \mu_{l_2}, \ldots, \mu_{l_{n_l}})'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \boldsymbol{\eta}_2', \ldots, \boldsymbol{\eta}_L')'$, $\mathbf{X}$ and $\boldsymbol{v}$ are defined similarly. Moreover, denote $N = \sum_{l=1}^L n_l T_l$, $Z_\mu = \text{diag}_L\{Z_{l\mu}\}$ with $Z_{l\mu} = I_{n_l} \otimes \boldsymbol{\iota}_{T_l}$, and $Z_\eta = \text{diag}_L\{Z_{l\eta}\}$ with $Z_{l\eta} = \boldsymbol{\iota}_{n_l} \otimes I_{T_l}$, where the symbol "$\otimes$" is the Kronecker product operator and I_k is an identity matrix of dimension k. From (3) and (4), we can show that

$$Q\mathbf{y} = Q\mathbf{X}\beta + Q\boldsymbol{v}, \tag{5}$$

where $Q = Q_{Z_\mu}Q_{Z_\eta}$, $Q_{Z_\mu} = \text{diag}_L\{I_{n_l} \otimes Q_{T_l}'\}$ and $Q_{Z_\eta} = \text{diag}_L\{P_{n_l} \otimes I_{T_l}\}$. Based on model (5), we can obtain a robust ordinary least squares estimator of β as follows,

$$\begin{aligned}\widehat{\beta} &= (\mathbf{X}'\text{diag}_L\{P_{n_l} \otimes P_{T_l}\}\mathbf{X})^{-1}\mathbf{X}'\text{diag}_L\{P_{n_l} \otimes P_{T_l}\}\mathbf{y} \\ &= (\sum_{l=1}^L\sum_{i=1}^{n_l}\widetilde{\mathbf{X}}_{l_i}'P_{T_l}\widetilde{\mathbf{X}}_{l_i})^{-1}\sum_{l=1}^L\sum_{i=1}^{n_l}\widetilde{\mathbf{X}}_{l_i}'P_{T_l}\widetilde{\mathbf{y}}_{l_i},\end{aligned} \tag{6}$$

where $P_k = I_k - \frac{1}{k}J_k$ with J_k denoting a $k \times k$ matrix of ones. Under some mild assumptions, we can show that

$$\sqrt{n}(\widehat{\beta} - \beta) \Longrightarrow N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}),$$

where $\Sigma_1 = \sum_{l=1}^L m_l[\mathbb{E}(\mathbf{X}_{l_i}'P_{T_l}\mathbf{X}_{l_i}) - \mathbb{E}\mathbf{X}_{l_i}'P_{T_l}\mathbb{E}\mathbf{X}_{l_i}]$ and $\Sigma_2 = \sum_{l=1}^L m_l\mathbb{E}[(\mathbf{X}_{l_i} - \mathbb{E}\mathbf{X}_{l_i})'P_{T_l}\boldsymbol{v}_{l_i}\boldsymbol{v}_{l_i}'(\mathbf{X}_{l_i} - \mathbb{E}\mathbf{X}_{l_i})]$ with $m_l = \lim_{n\to\infty}\frac{n_l}{n}$. The asymptotic results in this paper are based on the setting with large individual number and short time length. Besides we assume that $m_l = \lim_{n\to\infty}\frac{n_l}{n}$, which is a commonly used setting in the literature (see, e.g. Shao et al., 2011; Chowdhury, 1991).

3. Testing for individual effect

In this section, we consider to construct a statistic to test for individual effect in model (1). The hypotheses of individual effect test can be formalized as

$$H_0^\mu: \ \sigma_\mu^2 = 0 \quad \text{vs} \quad H_1^\mu: \ \sigma_\mu^2 > 0, \tag{7}$$

where σ_μ^2 is the variance of individual effect $\mu_{l_i}, l = 1, 2, \ldots, L$, $i = 1, 2, \ldots, n_l$. Denote the j-th column vector of matrix Q_{T_l} by $q_{l_j} = (q_{l_j1}, q_{l_j2}, \ldots, q_{l_jT_l})'$, $l = 1, 2, \ldots, L, j = 1, 2, \ldots, T_l - 1$, and then we have

$$\sum_{l=1}^L\sum_{i=1}^{n_l}\mathbb{E}\|Q_{T_l}'\widetilde{\mathbf{v}}_{l_i}\|^2 = \sum_{l=1}^L\sum_{i=1}^{n_l}\sum_{j=1}^{T_l-1}\mathbb{E}(q_{l_j}'\widetilde{\mathbf{v}}_{l_i})^2 = c_1\sigma_v^2,$$